

Opening Statistics

Murray A. Cameron

CSIRO and University of Technology Sydney, Sydney, Australia

E-mail: Murray.Cameron@CSIRO.au

Statistical work depends on access to software, relevant data, reports and publications, and colleagues and collaborators. Just as the Internet has transformed the business model of books and music, it promises (or threatens) to transform (statistical) research also. For example, when software is published electronically it is rapidly used by others, though that makes the potential impact of bugs in the software much greater. Developments in email, blogs and social media have created possibilities for more widespread “collaborations” that have fewer overheads and the potential to advance research far more quickly. This is the thesis of Nielsen (2011) and others. The papers here, that were presented in a “late-breaking session” at the ISI’s World Statistics Congress in 2013 (session IPS108), look at various aspects of openness as they relate to statistics - some of the opportunities and the potential pitfalls and blockages. The papers created lively discussion from the audience during the session, though the topic and the issues are far broader than could be covered in a couple of hours.

In the first paper, Victoria Stodden (Columbia University) describes standards recommended by meetings of researchers seeking to ensure that computational research is published with sufficient detail available so that others working in the field can replicate the results. They recommend actions by researchers, granting agencies and research journals as well as employers and educators.

The other three papers in this collection consider different sorts of data. Ron Sandland (ANDS) describes the infrastructure investments being made in Australia to facilitate the management and sharing of very large

scientific data sets for collaboration in fields such as astronomy, bioinformatics, climate research and environmental science. He also flags the opportunities available for statisticians that are not being grasped. Alan Karr (NISS) highlights many of the difficulties of sharing data. He argues that data availability is a *public good* and recognising this highlights the complexities of making data widely available as well as suggesting paths to solution. As with other public goods, there is a cost to the individual (the data publisher) and at least some of that cost should be compensated by the wider community of beneficiaries. The final paper, by Misha Belkindas and Eric Swanson illustrates some of the issues mentioned in Alan Karr’s paper. In it, they consider the potential difficulties faced by governments in making data open. They cite legal, organisational, technical and political difficulties as well as the financial cost and, while recognising the contribution of foundations and international agencies, they argue that non-governmental organisations also have a role to play.

In summary, the papers highlight that, while openness offers many benefits, the practice will require the development of standards of practice, tools to make compliance easy, education to make them understood and culture change to ensure that the benefits are achieved.

Reference

- [1] M. Nielsen, *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press, 2011.